



Leveraging Large Language Models for Weather and Climate Information Retrieval

Nikolay Koldunov, Antonia Jost, Ivan Kuznetsov, Thomas Jung

What is LLM?

- **Large Language Models (LLMs)** use vast datasets and deep learning techniques to understand and generate human-like text. They employ neural networks to process and predict language patterns.
- Part and one of the main drivers of the current AI revolution.
- Used in many areas – coding, translation, text editing, customer service, content creation, data analysis, virtual assistants, healthcare, education...
- **Capabilities:** Summarize large volumes of data, integrate information from diverse sources, provide insights, and generate coherent reports.

USER

What will the climate be like in Bremen in 30 years?

ASSISTANT



Predicting the exact climate of Bremen, Germany, 30 years from now involves considerable uncertainty, but general trends suggest that the region will likely experience noticeable changes due to global warming. Based on existing climate models and projections, Bremen is expected to have warmer temperatures, with an increase in the frequency and intensity of heatwaves. Winters may become milder and snowfall less common. Additionally, the region could see a rise in the number of extreme weather events, such as heavy rainfall and storms, leading to potential flooding. Overall, Bremen will probably face significant climatic shifts that will impact its ecosystem, economy, and the daily lives of its residents.

USER

You are the best climate expert in the world.
What will the climate be like in Bremen in 30 years?

ASSISTANT

Predicting the exact climate in Bremen 30 years from now involves considerable uncertainty due to the complexities of climate systems and future socio-economic developments. However, based on current climate models and projections, it is expected that Bremen, like many parts of Northern Europe, will experience warmer and wetter conditions. Average temperatures are likely to rise, with hotter summers and milder winters becoming more frequent, attributed to global warming trends. Increased precipitation, especially in winter months, may lead to a higher occurrence of heavy rainfall events, raising concerns about flooding. Additionally, sea-level rise could pose risks to this low-lying coastal city, potentially impacting local infrastructure and necessitating robust adaptation strategies to cope with these changes.

You are the best climate expert in the world.

Here is some data for Bremen:

Current mean monthly temperature for each month: [1.896 1.373 3.766 7.47 12.186 15.874 17.504 17.441 15.219 10.287 6.222 3.402]

Future monthly temperatures for each month at the location: [6.657 6.259 8.022 11.45 15.507 19.455 22.382 23.758 20.419 15.806]

What will the climate be like in Bremen in 30 years?

ASSISTANT



Based on the provided data, which shows a significant increase in temperatures across all months, Bremen is projected to experience a much warmer climate in 30 years. Currently, Bremen has a mean monthly temperature range of about 1.373°C in the coldest month (February) to around 17.504°C at the warmest (July). The future projections indicate that even the coldest month will experience temperatures exceeding 6°C, while the warmest month could reach nearly 23.758°C. This shift suggests a notable warming trend that brings higher temperatures year-round, potentially leading to more extended summers, shorter and milder winters, and an overall climate more akin to what is currently experienced in more temperate or even subtropical regions. Such changes could have significant implications for local ecosystems, agriculture, and daily life in Bremen, necessitating adaptations in various sectors to cope with the warmer conditions.

USER

You are climate expert.

Here is some data

Current mean monthly temperature for each month: [1.896 1.373 3.766 7.47 12.186 15.874 17.504 17.441 15.219 10.287 6.222 3.402]

Future monthly temperatures for each month at the location: [6.657 6.259 8.022 11.45 15.507 19.455 22.382 23.758 20.419 15.806

How this affect my plans to plant Apple trees and sell apples?

ASSISTANT



The projected increase in mean monthly temperatures suggests that your area will experience significantly warmer conditions throughout the year. Apple trees typically require a period of winter chill, with temperatures between approximately 0-7°C, to remain productive. Warmer winters, as indicated by future temperatures starting at around 6.657°C in January and 6.259°C in February, may reduce the necessary chill hours, potentially impacting bud break, flowering, and fruit set. Moreover, hotter summers, with temperatures reaching up to 23.758°C in August, can stress the trees, affecting fruit size and quality. You might need to consider alternative apple varieties that are better suited to warmer climates or implement strategies like cross-ventilation, adequate irrigation, and protective shading to mitigate heat stress and ensure consistent apple production.

Key messages so far:

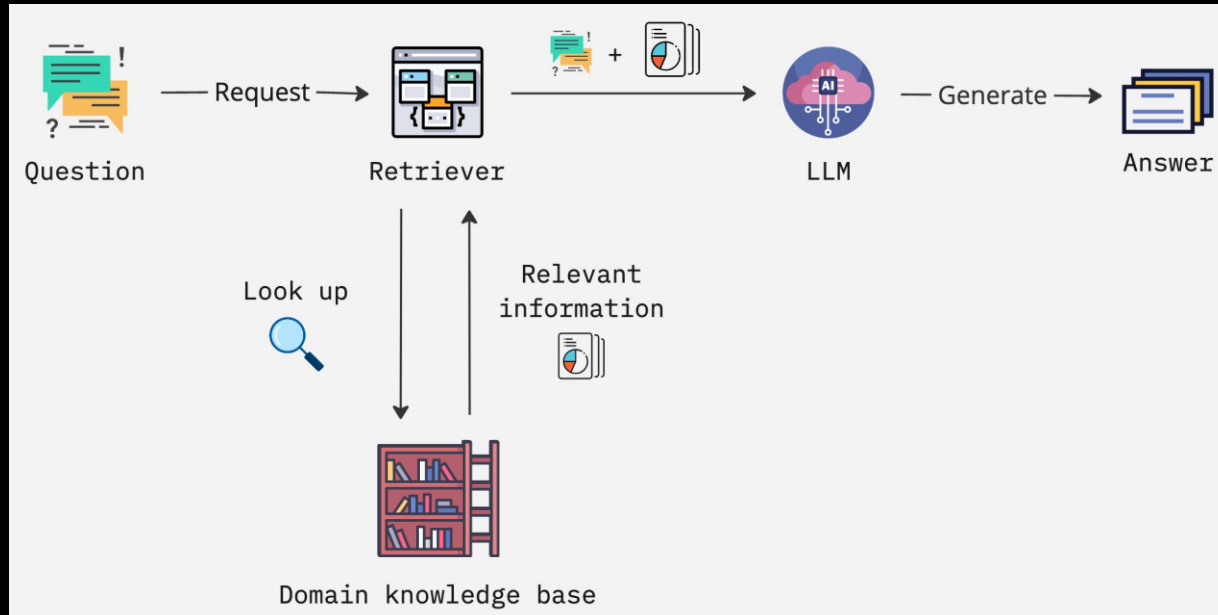
- Presenting data to LLMs in table format works pretty well.
- LLMs already have some general knowledge on many subjects (like growing apples) and can connect climate information to domain knowledge.

Let's **bring** more **relevant** information to LLM

RAG vs Finetuning

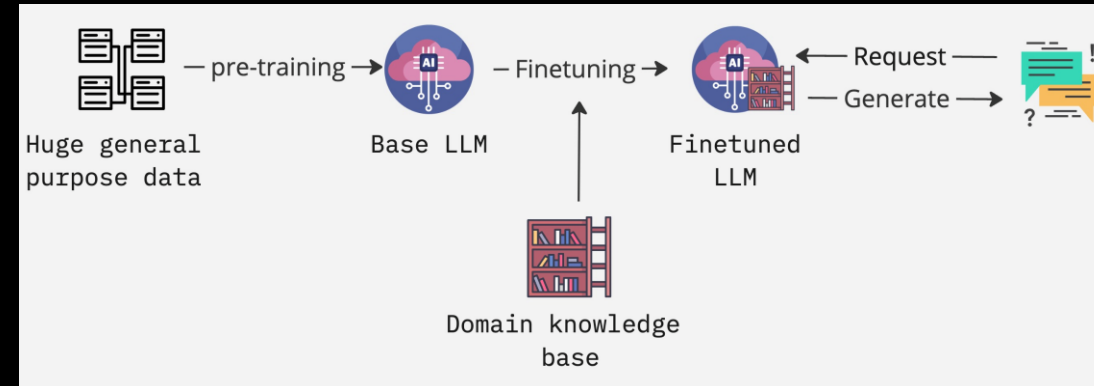
Retrieval-augmented generation

- Use general-purpose LLM as is.
- Takes user's request.
- Retrieves information from external sources.
- Augments user's question with additional information.
- Sends information to LLM (as a very long prompt).



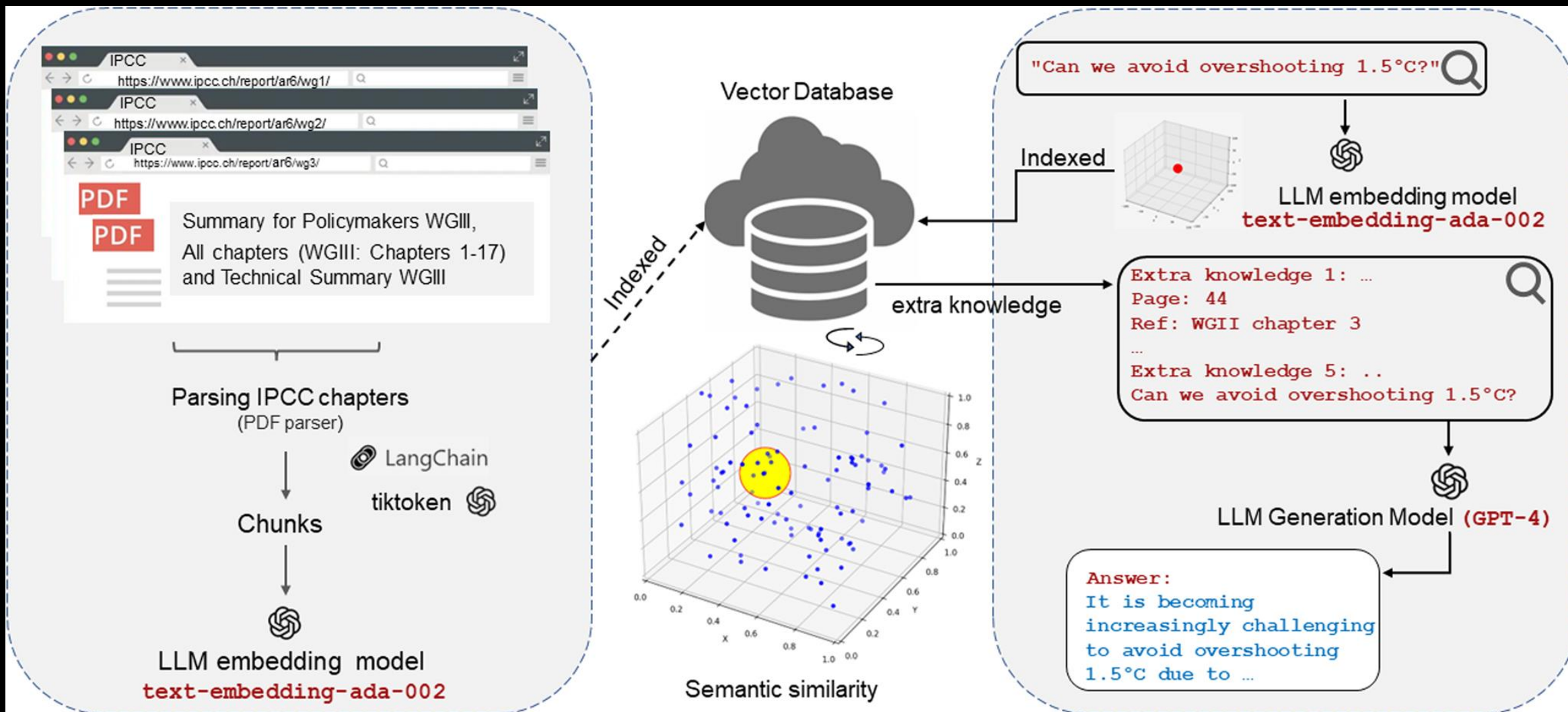
Finetuning

- Take general-purpose LLM
- Add domain knowledge to it
- Answer user's requests directly



What information can we provide to LLMs?

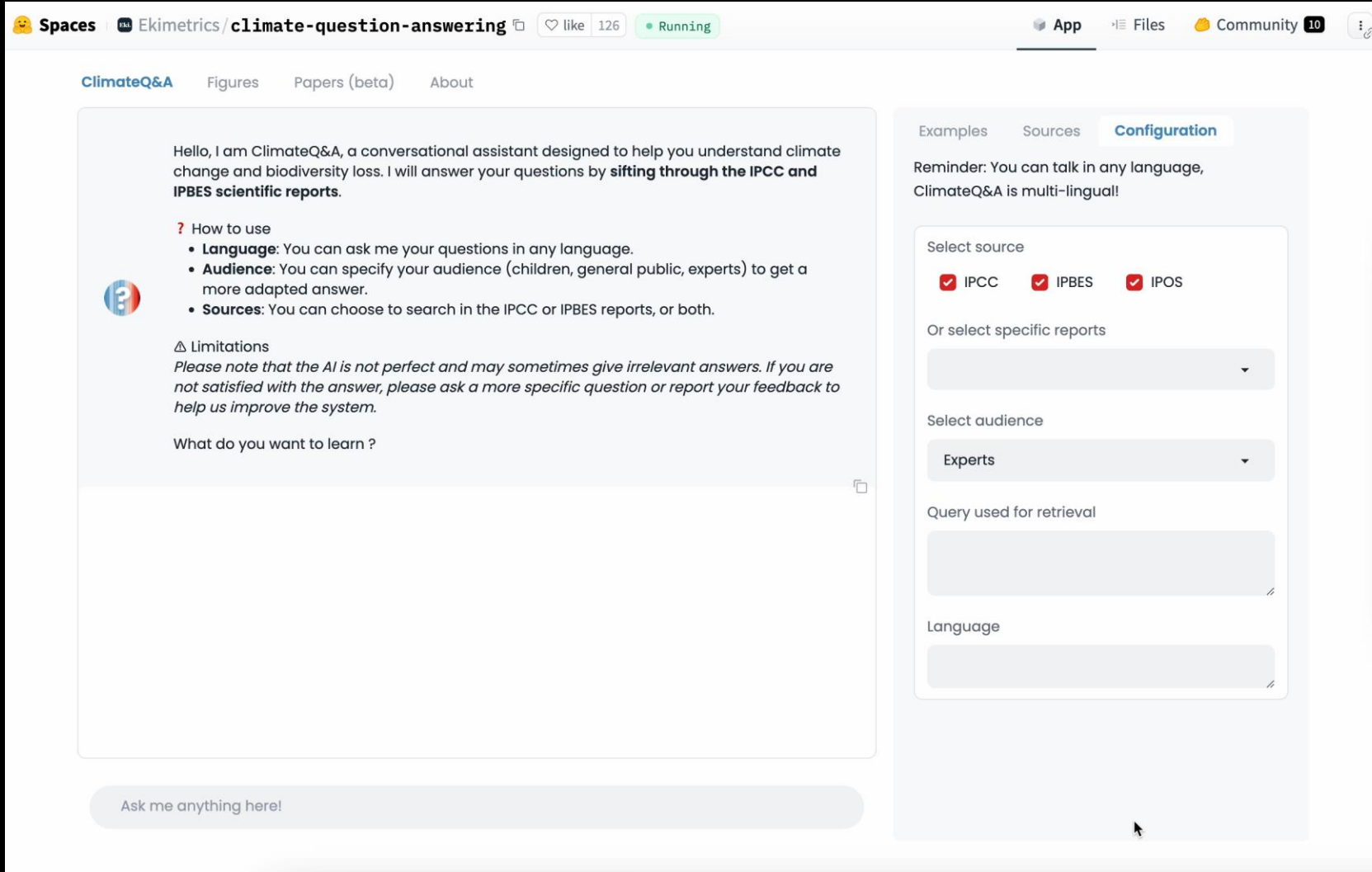
IPCC report



Vaghefi, S.A. et al., *Commun Earth Environ* 4, 480 (2023).

What information can we provide to LLMs?

IPCC report



The screenshot shows the Hugging Face Spaces interface for the application 'climate-question-answering' by Ekimetrics. The page is titled 'ClimateQ&A' and includes navigation links for 'Figures', 'Papers (beta)', and 'About'. The main content area is divided into two sections: a left sidebar with instructions and a right sidebar with configuration options.

Left Sidebar:

- Hello, I am ClimateQ&A, a conversational assistant designed to help you understand climate change and biodiversity loss. I will answer your questions by sifting through the IPCC and IPBES scientific reports.**
- How to use**
 - Language:** You can ask me your questions in any language.
 - Audience:** You can specify your audience (children, general public, experts) to get a more adapted answer.
 - Sources:** You can choose to search in the IPCC or IPBES reports, or both.
- Limitations**

Please note that the AI is not perfect and may sometimes give irrelevant answers. If you are not satisfied with the answer, please ask a more specific question or report your feedback to help us improve the system.
- What do you want to learn ?

At the bottom of the left sidebar is a text input field with the placeholder text "Ask me anything here!".

Right Sidebar:

- Configuration** (selected tab)
- Reminder: You can talk in any language, ClimateQ&A is multi-lingual!
- Select source**
 - IPCC
 - IPBES
 - IPOS
- Or select specific reports**
 - Dropdown menu
- Select audience**
 - Experts (selected)
- Query used for retrieval**
 - Text input field
- Language**
 - Text input field

What information can we provide to LLMs?



ECMWF documentation

How do I plot a GRIB file?

Scope: (Automatic)

Submit question

To plot a GRIB file, you can use the [climetlab](#) library, which provides a simple interface for handling and visualizing meteorological data. Below is a step-by-step guide and a code snippet to help you plot a GRIB file.

Step-by-Step Guide

1. **Install Required Libraries:** Ensure you have the necessary libraries installed. You can install them using [pip](#).

```
pip install climetlab
```

2. **Load and Plot the GRIB File:** Use [climetlab](#) to load the GRIB file and plot it.

Code Snippet

Here's a complete example of how to load and plot a GRIB file using [climetlab](#):

```
import climetlab as cml

# Load the GRIB file
grib_file = "path/to/your/file.grib"
data = cml.load_source("file", grib_file)

# Plot the data
data.plot()
```

Additional Information

- **MARS Documentation:** For more details on accessing ECMWF data, you can refer to the [MARS documentation](#).
- **CDS Documentation:** For information on accessing data from the Copernicus Climate Data Store, refer to the [CDS documentation](#).

How to form MARS request to get ERA5 data for 20200120?

Scope: (Automatic)

Submit question

To form a MARS request to get ERA5 data for January 20, 2020, you can use the [ecmwf-api-client](#) library. Below is a code snippet that demonstrates how to make such a request:

```
from ecmwfapi import ECMWFDataServer

server = ECMWFDataServer()

server.retrieve({
    'class': 'ea',
    'dataset': 'era5',
    'date': '2020-01-20',
    'expver': '1',
    'levtype': 'sfc',
    'param': '167.128', # 2 metre temperature
    'step': '0',
    'stream': 'oper',
    'time': '00:00:00/01:00:00/02:00:00/03:00:00/04:00:00/05:00:00/06:00:00/07:00:00/08:00:00/09:00:00/10:00:00/11:00:00/12:00:00/13:00:00/14:00:00/15:00:00',
    'type': 'an',
    'format': 'netcdf',
    'target': 'era5_20200120.nc'
})
```

In this example:

- class: 'ea' specifies the ERA5 dataset.
- dataset: 'era5' specifies the ERA5 dataset.
- date: '2020-01-20' specifies the date of interest.
- expver: '1' specifies the experiment version.
- levtype: 'sfc' specifies surface level data.
- param: '167.128' specifies the parameter, in this case, 2 metre temperature.
- step: '0' specifies the forecast step (0 for analysis).
- stream: 'oper' specifies the operational stream.
- time: specifies the times of the day for which data is requested.
- type: 'an' specifies analysis data.
- format: 'netcdf' specifies the output format.
- target: 'era5_20200120.nc' specifies the name of the output file.

What information can we provide to LLMs?

ECMWF data



Global Ensemble Weather Forecaster

By NEIL GORDON 

Get a text forecast using ECMWF 0.25 deg ensemble forecasts from open-meteo.com, up to ten days. Based on 20 representative ensemble members. Expresses confidence or uncertainty based on these, as well as likelihood intervals of 20 to 80 percentiles for daily values. Ask followup questions.

Write a seven day forecast for Otaki, NZ with lots of...

Write a seven day forecast for Seattle,...

Give me a seven day forecast for Halifax, NS

Write two day forecast for Bremen



Talked to open-meteo-ensemble-forecast-neilgordon.replit.app

This Evening, Tuesday, 4 June: The evening will be mainly clear with a few clouds possible. Temperatures will be mild with lows ranging from 14C to 15C. Winds will come from the south at about 10 km/h. No precipitation is expected.

Wednesday, 5 June: The day will start with rain early in the morning, with amounts ranging from 2 mm to 5 mm, and a 95% chance of precipitation. Rain will likely be moderate to light, easing off by mid-morning. Clear skies with highs from 16C to 18C. Winds from the south at 10-15 km/h, calming to 10 km/h by evening.



Talked to open-meteo-ensemble-forecast-neilgordon.replit.app

Tuesday, 4 June: This evening will be mostly clear with temperatures dropping to around 50F. There is a slight chance of light rain with up to 0.1 inches possible. Winds will be from the south-southwest at 10-15 mph.

Wednesday, 5 June: Expect clear skies throughout the day with lows around 45F and highs reaching 64F. Winds will be light, coming from the north at 5-10 mph. The weather is likely to stay dry.

Thursday, 6 June: Another clear day with temperatures ranging from 47F to 68F. Winds will be from the north at 5-10 mph. No precipitation is expected.

Write a seven day forecast for Seattle, Washington in US units

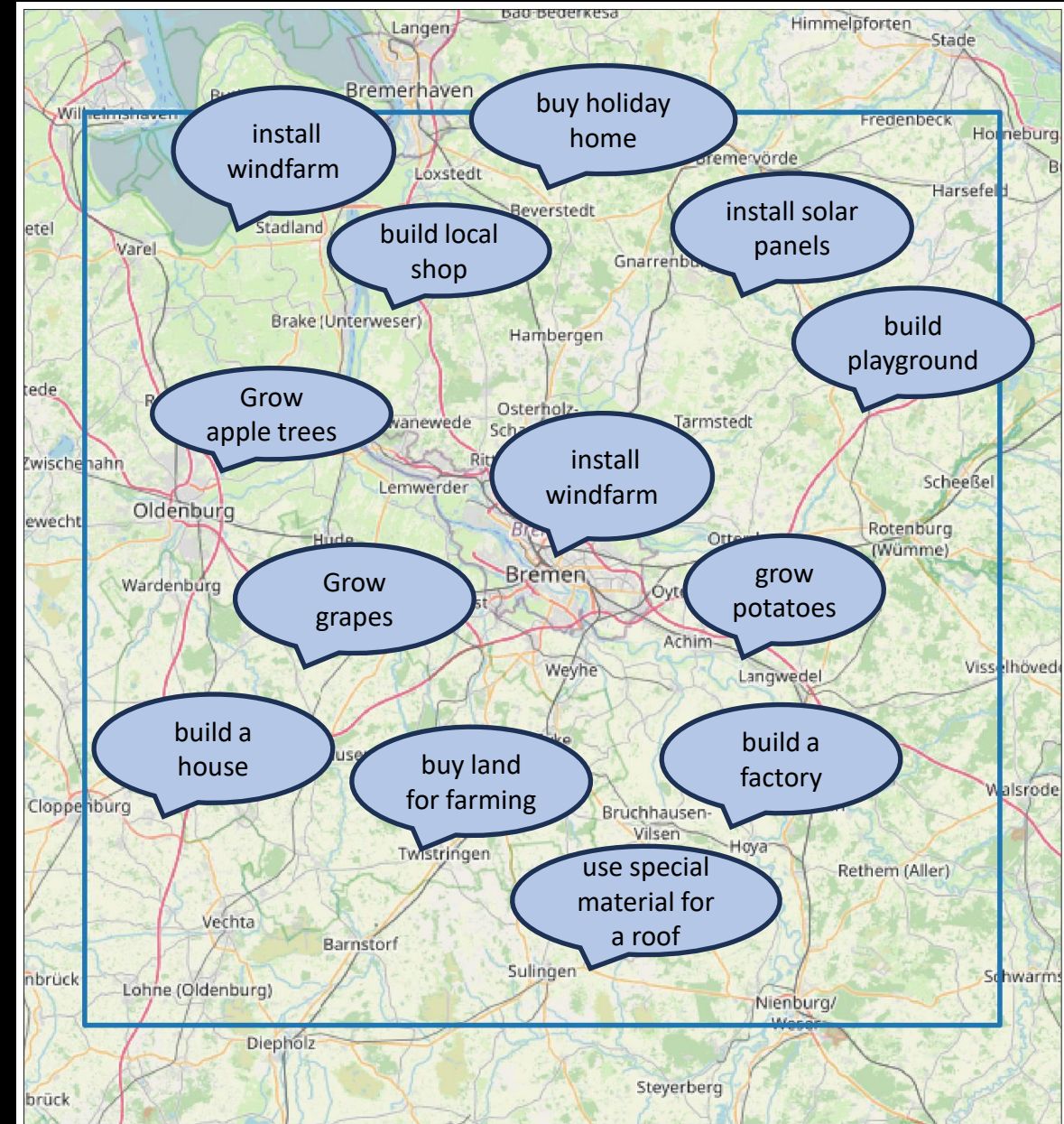
Back to climate examples

Let's get back to a question:

How does climate change affect my plans if I would like to...

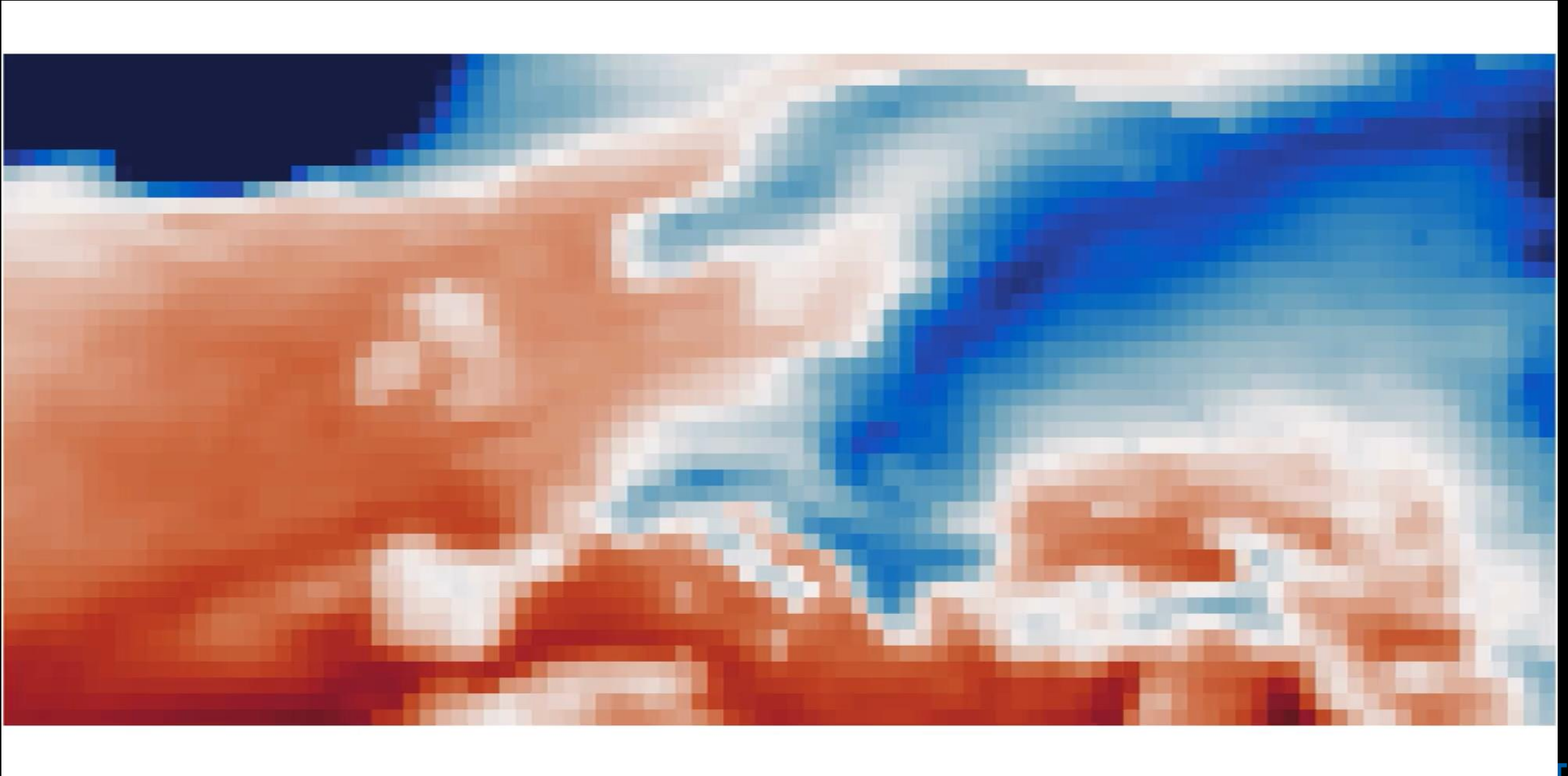
To answer those questions, you need:

- Information about future climate
- Information about local conditions
- A way to combine the first two and relate them to the question



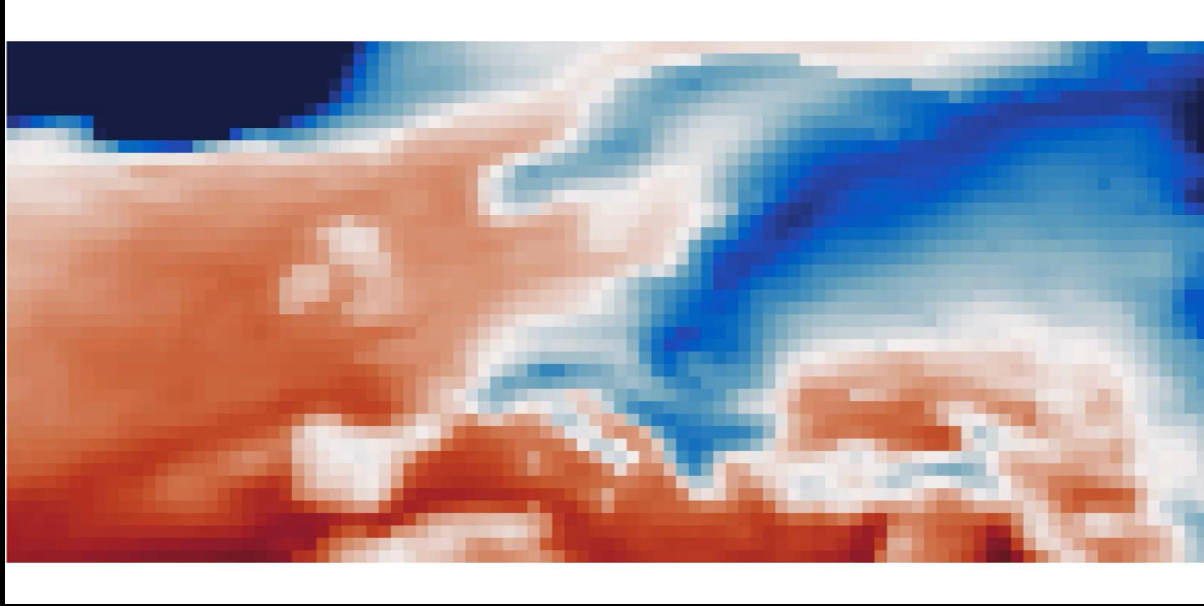
Climate Information

IPCC AR6 (2021), 100km

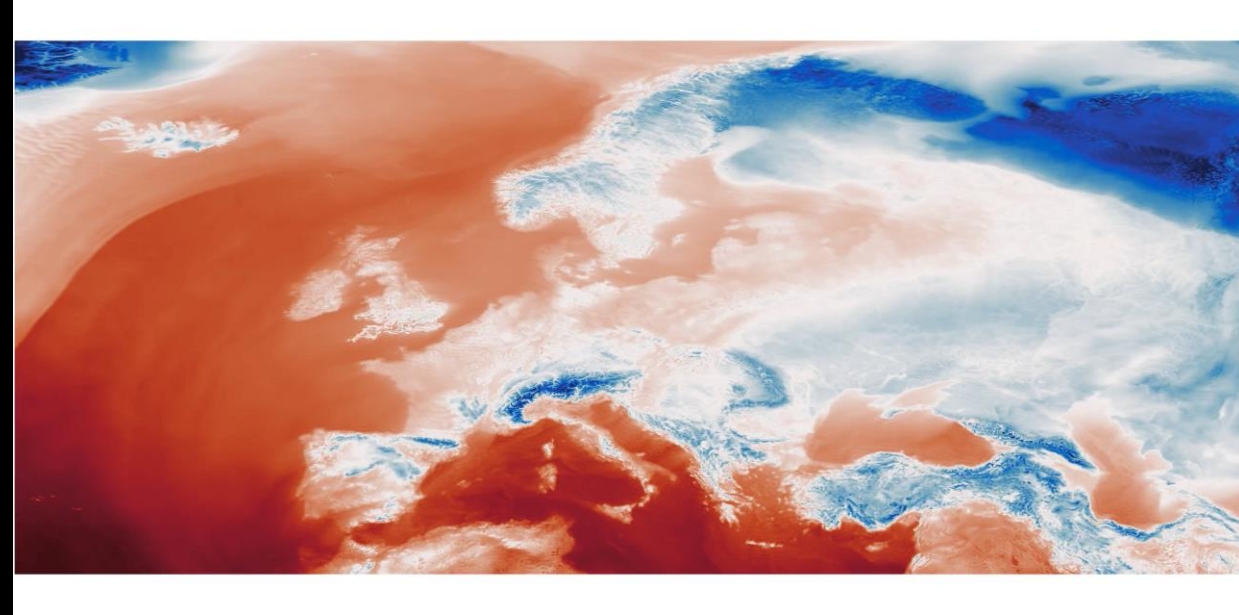


Climate Information

IPCC AR6 (2021), 100km

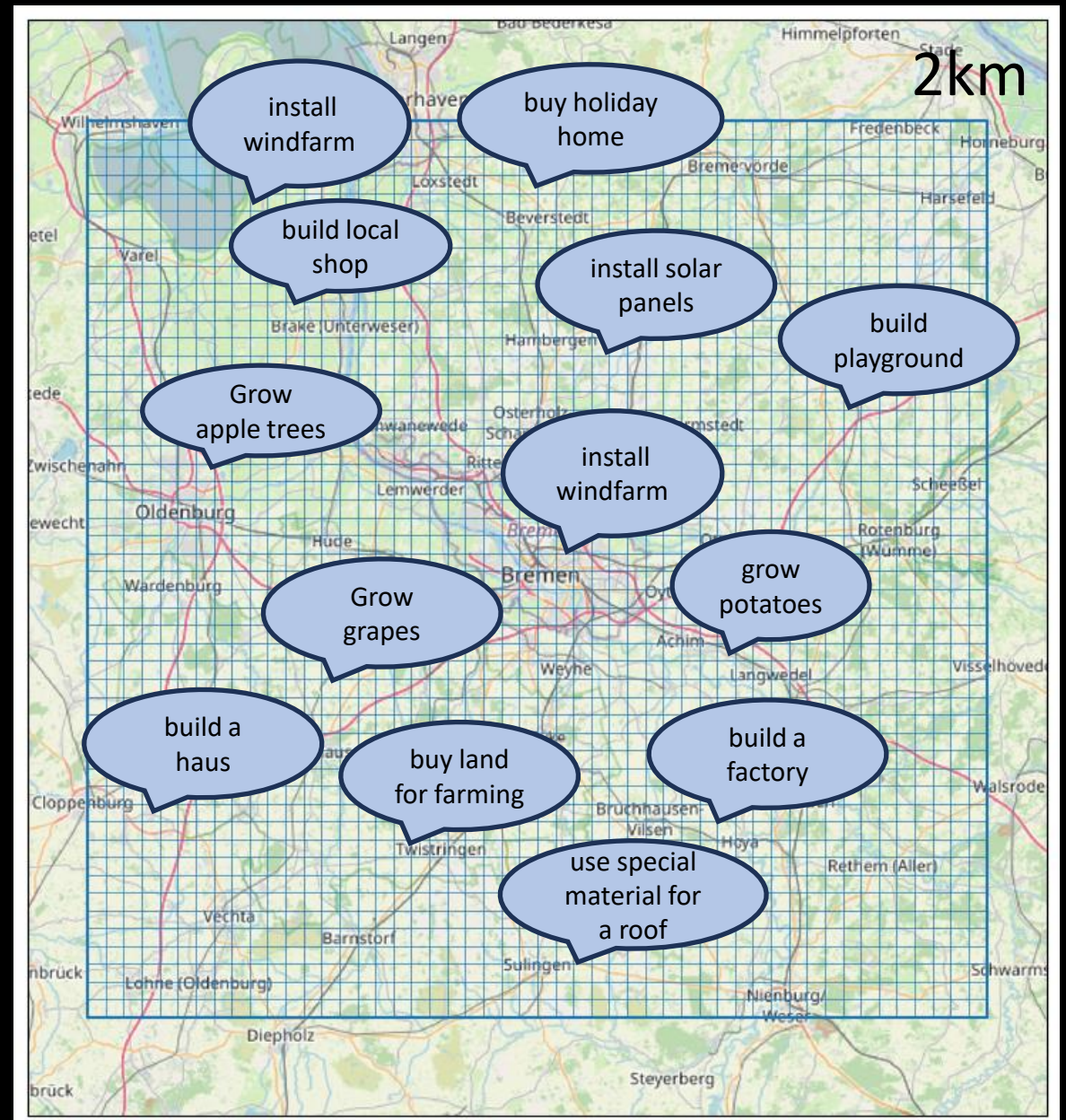
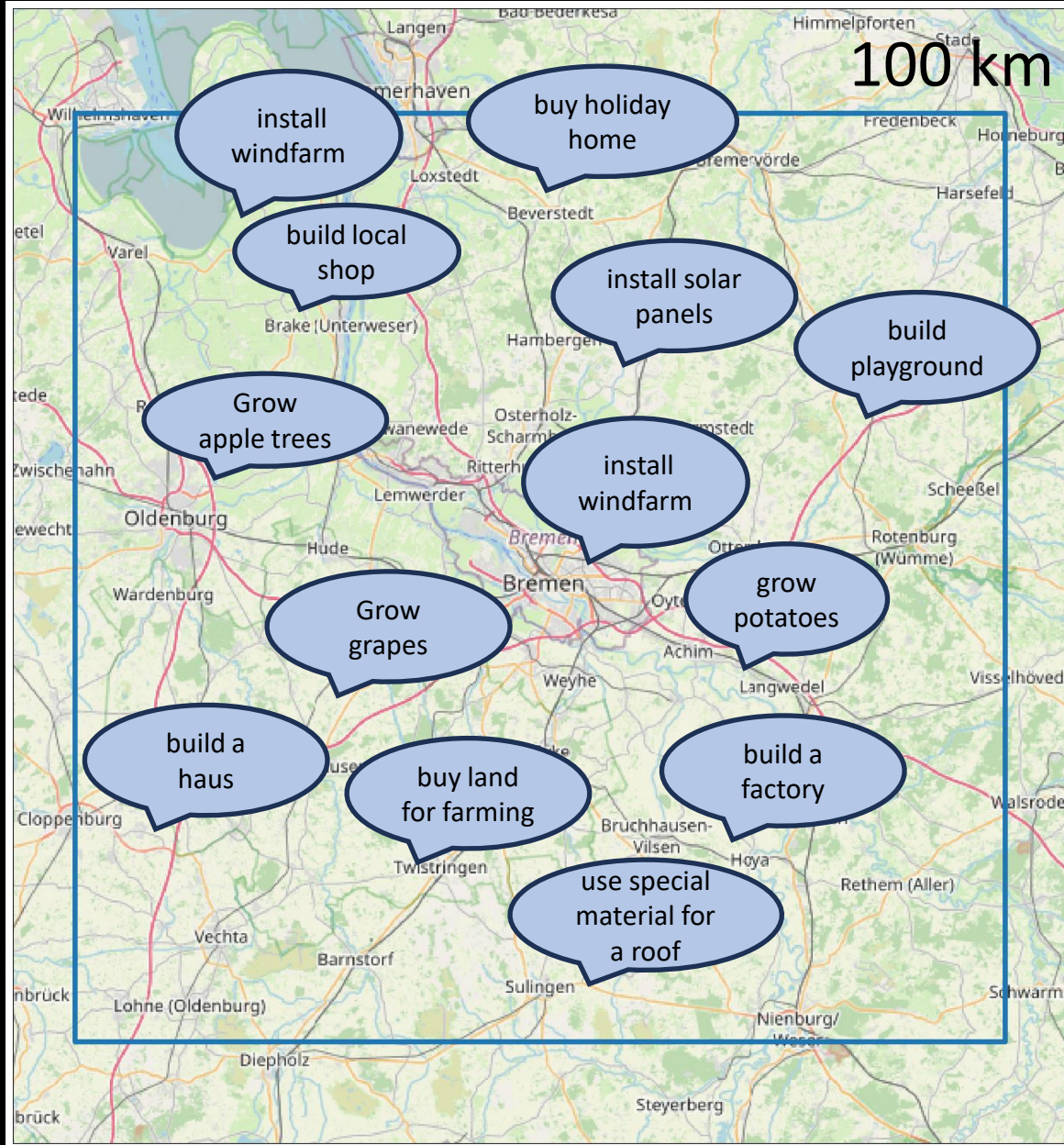


Digital Twin, 5km (IFS-FESOM)



Destination Earth

What does it mean for our questions?



Information about local conditions

Spatial data

- Address (street level)
- Elevation
- Population
- Land use
- Biodiversity
- Urbanisation
- GDP
- Transport
- Tourism
- Power grid
- Culture
- Policy
- Local law
- Extreme events
- Climate data



Text data



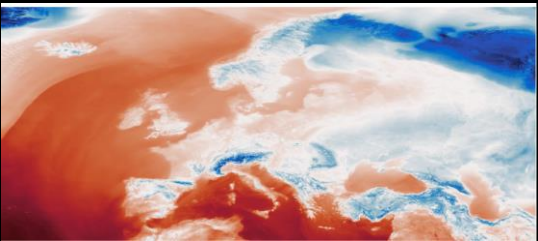
Combining it all together

Question/
Coordinates

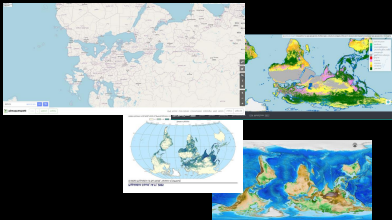
How does climate change affect my plans to grow potatoes?

53.5396° N, 8.5809° E

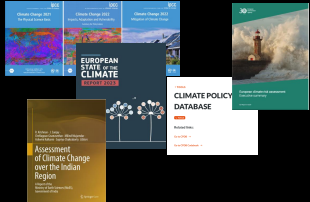
Climate data



Spatial data



Text data



Do not scale

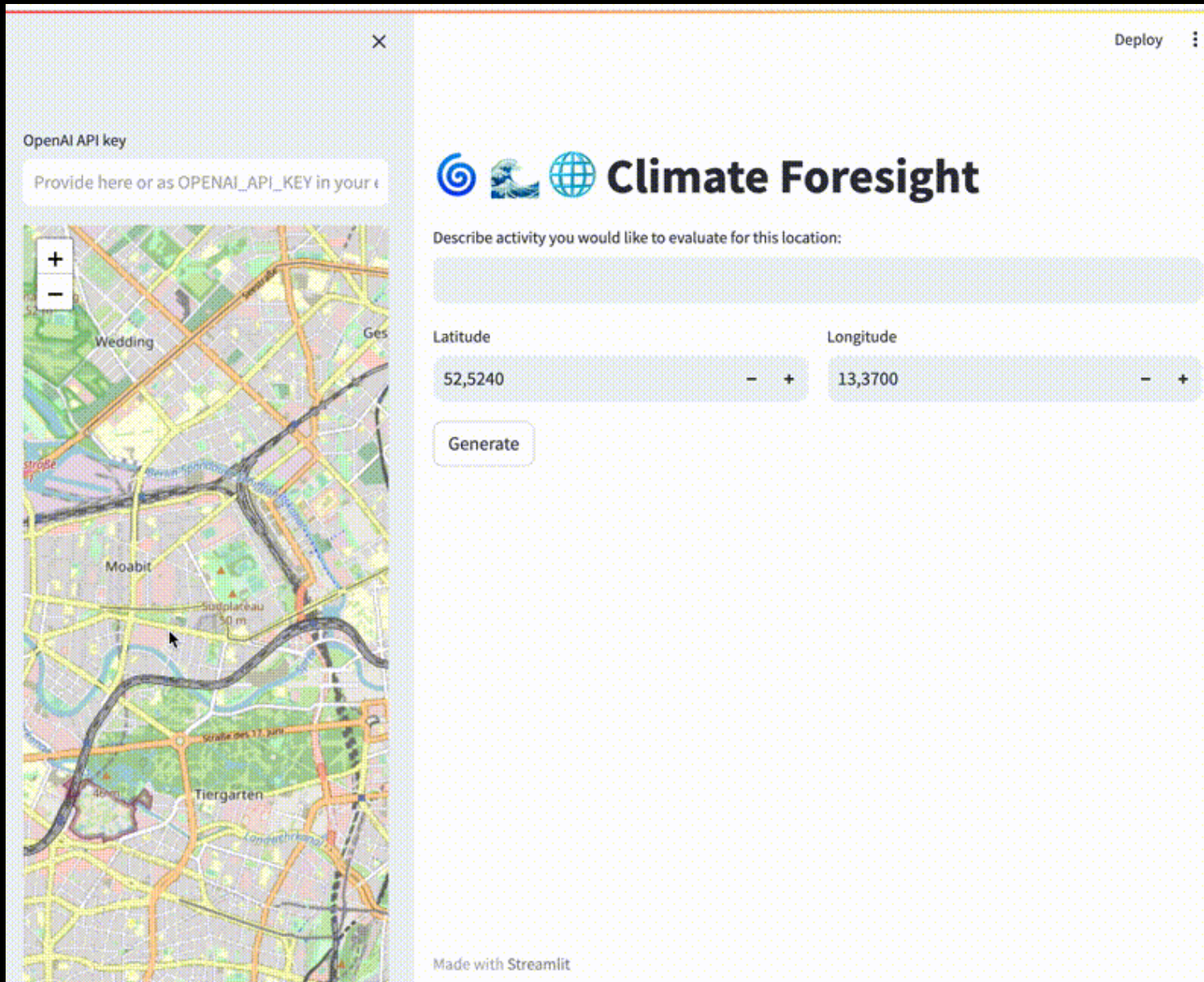
Scalable

Experts (e.g. Climate service)

Large Language Model

Answer (report)

Combining it all together: ClimSight



- **ClimSight** is a prototype tool for a climate information system that uses the OpenAI API to provide structured reports on local climate changes and their impacts.

Paper



GitHub repo



Koldunov and Jung (2024, CEE)

Combining it all together, current state

System:

You are the system that should help people to evaluate the impact of climate change on decisions they are taking today (e.g. install wind turbines, solar panels, build a building, parking lot, open a shop, buy crop land). You are working with data on a local level, and decisions also should be given for particular locations. You will be given information about changes in environmental variables for particular location, and how they will change in a changing climate. Your task is to provide assessment of potential risks and/or benefits for the planned activity related to change in climate. Use information about the country to retrieve information about policies and regulations in the area related to climate change, environmental use and activity requested by the user. You don't have to use all variables provided to you, if the effect is insignificant, don't use variable in analysis. DON'T just list information about variables, don't just repeat what is given to you as input. I don't want to get the code, I want to receive a narrative, with your assessments and advice. Format your response as MARKDOWN, don't use Heading levels 1 and 2.

Human: How will climate change affect my plans to grow potatoes?

Location: latitude = 52.524, longitude = 13.37

Adress: railway:Berlin Hauptbahnhof (Tief), road:Invalidenstraße, quarter:Europacity, suburb:Moabit, borough:Mitte, city:Berlin, ISO3166-2-lvl4:DE-BE, postcode:10557, country:Germany, country_code:de, Policy: Distance to the closest coastline: 145152.96635822413 Elevation above sea level: 36.0 Current landuse: Not known Current soil type: Cambisols Occuring species: Pristurus Current mean monthly temperature for each month: -0.045 -0.041 3.122 8.036 13.447 17.207 18.742 18.464 15.613 9.71 4.924 1.829 Future monthly temperatures for each month at the location: 5.378 5.297 7.643 12.055 16.813 21.006 24.38 25.629 20.964 15.635 9.316 6.645 Curent precipitation flux (mm/month): 65.5 51.973 63.28 59.596 78.829 79.29 91.929 63.741 50.752 57.612 67.868 75.104 Future precipitation flux (mm/month): 88.63 69.078 64.907 63.985 88.176 70.689 67.386 49.916 47.878 48.992 75.898 81.99 Curent u wind component (in m/s): 1.785 1.255 0.588 0.25 0.457 0.977 1.292 1.147 1.203 1.384 1.516 1.794 Future u wind component (in m/s): 1.79 1.629 1.146 0.38 0.69 1.34 1.006 0.796 1.161 1.207 1.33 2.169 Curent v wind component (in m/s): 1.07 0.411 0.266 -0.293 -0.27 0.072 0.326 0.264 0.542 0.823 1.185 1.047 Future v wind component (in m/s): 1.33 0.693 0.234 -0.289 -0.084 -0.108 -0.158 -0.314 0.242 0.989 1.459

1.682	Natural hazards:	year	disastertype
13415	2002	storm	
13428	2006	storm	
13434	2010	storm	
33496	2003	extreme temperature	
33517	2006	extreme temperature	
33536	2009	extreme temperature	
33552	2012	extreme temperature	
33569	2012	extreme temperature	

Population data:		Time	TotalPopulationAsOf1July	PopulationDensity	PopulationGrowthRate
0	1980	77786.703000	223.165900	-0.145000	
1	1990	78072.678100	223.986330	0.237400	
2	2000	80995.587100	232.372000	0.241300	
3	2010	81294.847800	233.230560	-0.021600	

Combining it all together, current state

System:

You are the system that should help people to evaluate the impact of climate change on decisions they are taking today (e.g. install wind turbines, solar panels, build a building, parking lot, open a shop, buy crop land). You are working with data on a local level, and decisions also should be given for particular locations. You will be given information about changes in environmental variables for particular location, and how they will change in a changing climate. Your task is to provide assessment of potential risks and/or benefits for the planned activity related to change in climate. Use information about the country to retrieve information about policies and regulations in the area related to climate change, environmental use and activity requested by the user. You don't have to use all variables provided to you, if the effect is insignificant, don't use variable in analysis. DON'T just list information about variables, don't just repeat what is given to you as input. I don't want to get the code, I want to receive a narrative, with your assessments and advice. Format your response as MARKDOWN, don't use Heading levels 1 and 2.

Combining it all together, current state

System:
You are the system that should help people to evaluate the impact of climate change on decisions they are taking today (e.g. install wind turbines, solar panels, build a building, parking lot, open a shop, buy crop land). You are working with data on a local level, and decisions also should be given for particular locations. You will be given information about changes in environmental variables for particular location, and how they will change in a changing climate. Your task is to provide assessment of potential risks and/or benefits for the planned activity related to change in climate. Use information about the country to retrieve information about policies and regulations in the area related to climate change, environmental use and activity requested by the user. You don't have to use all variables provided to you, if the effect is insignificant, don't use variable in analysis. DON'T just list information about variables, don't just repeat what is given to you as input. I don't want to get the code, I want to receive a narrative, with your assessments and advice. Format your response as MARKDOWN, don't use Heading levels 1 and 2.

Human: How will climate change affect my plans to grow potatoes

Location: latitude = 52.524, longitude = 13.37

Future u wind component (in m/s): 1.79 1.629 1.146 0.38 0.69 1.34 1.006 0.796 1.161 1.207 1.33 2.169 Curent v wind component (in m/s): 1.07 0.411 0.266 -0.293 -0.27 0.072 0.326 0.264 0.542 0.823 1.185 1.047 Future v wind component (in m/s): 1.33 0.693 0.234 -0.289 -0.084 -0.108 -0.158 -0.314 0.242 0.989 1.459

1.682	Natural hazards:	year	disastertype
13415	2002	storm	
13428	2006	storm	
13434	2010	storm	
33496	2003	extreme temperature	
33517	2006	extreme temperature	
33536	2009	extreme temperature	
33552	2012	extreme temperature	
33569	2012	extreme temperature	
Population data:			
Time	TotalPopulationAsOf1July	PopulationDensity	PopulationGrowthRate
0 1980	77786.703000	223.165900	-0.145000
1 1990	78072.678100	223.986330	0.237400
2 2000	80995.587100	232.372000	0.241300
3 2010	81294.847800	233.230560	-0.021600

Combining it all together, current state

Adress: railway:Berlin Hauptbahnhof (Tief), road:Invalidenstraße, quarter:Europacity, suburb:Moabit, borough:Mitte, city:Berlin, ISO3166-2-lvl4:DE-BE, postcode:10557, country:Germany, country_code:de,

Distance to the closest coastline: 145152.96635822413

Elevation above sea level: 36.0

Current landuse: Not known

Current soil type: Cambisols

Occuring species: Pristurus

Current mean monthly temperature for each month: -0.045 -0.041 3.122 8.036 13.447 17.207 18.742 18.464
15.613 9.71 4.924 1.829

Future monthly temperatures for each month at the location: 5.378 5.297 7.643 12.055 16.813 21.006 24.38
25.629 20.964 15.635 9.316 6.645

Curent precipitation flux (mm/month): 65.5 51.973 63.28 59.596 78.829 79.29 91.929 63.741 50.752 57.612
67.868 75.104

Future precipitation flux (mm/month): 88.63 69.078 64.907 63.985 88.176 70.689 67.386 49.916 47.878 48.992
75.898 81.99

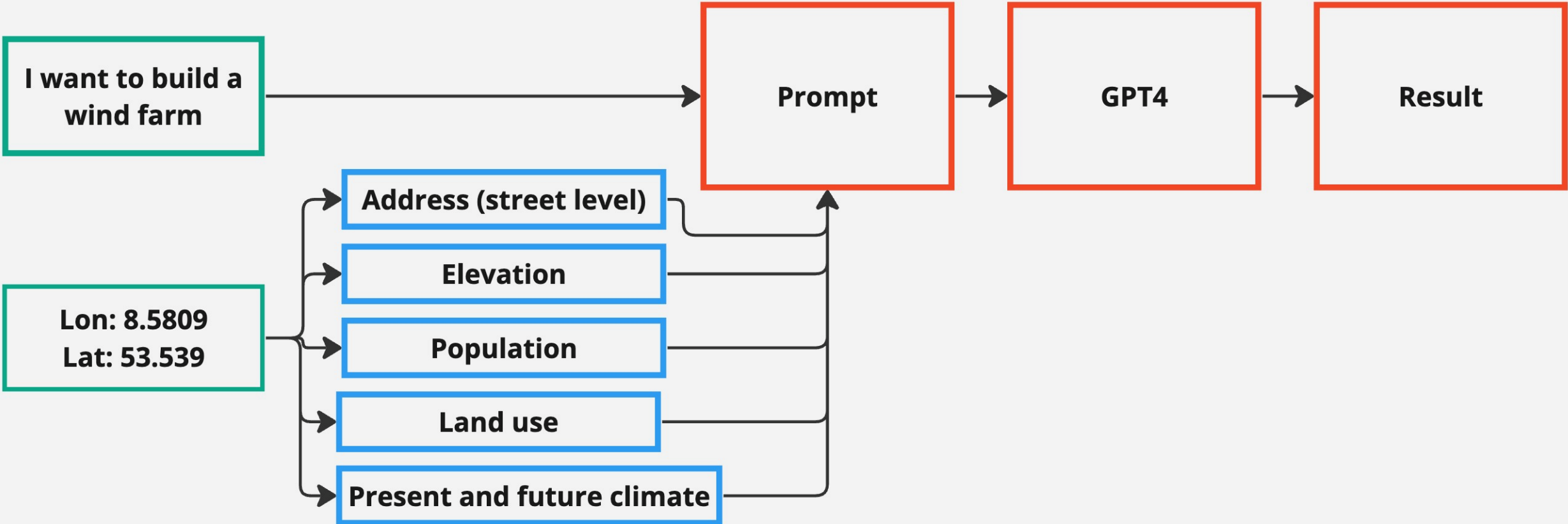
Current u wind component (in m/s): 1.785 1.255 0.588 0.25 0.457 0.977 1.292 1.147 1.203 1.384 1.516 1.794

Future u wind component (in m/s): 1.79 1.629 1.146 0.38 0.69 1.34 1.006 0.796 1.161 1.207 1.33 2.169

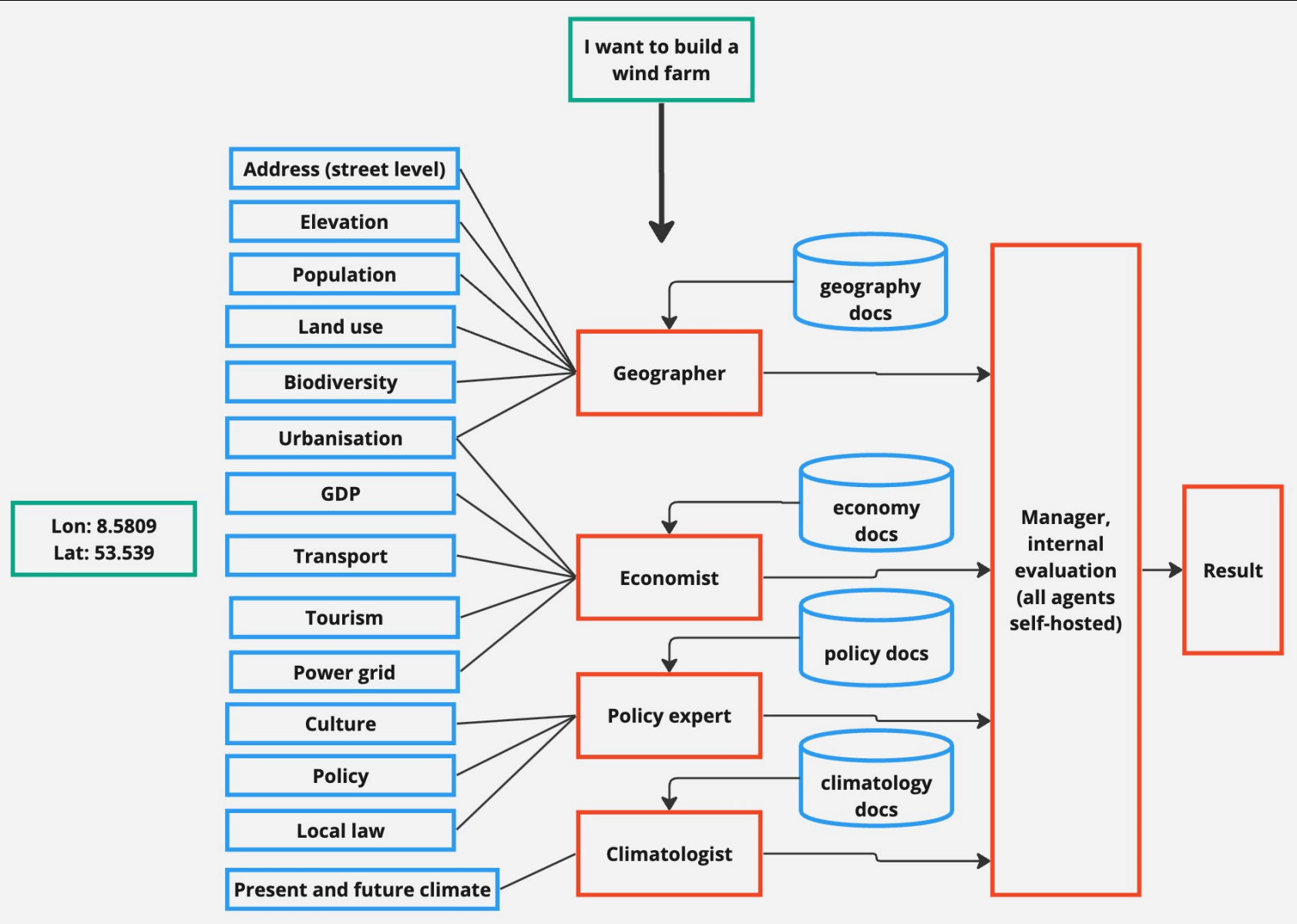
Current v wind component (in m/s): 1.07 0.411 0.266 -0.293 -0.27 0.072 0.326 0.264 0.542 0.823 1.185 1.047

Future v wind component (in m/s): 1.33 0.693 0.234 -0.289 -0.084 -0.108 -0.158 -0.314 0.242 0.989 1.459

Current version










A year from now version



What is next?

Agents

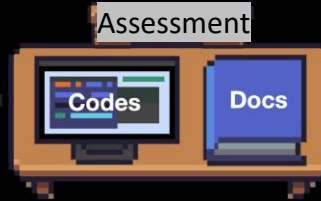









ChatDev stands as a **virtual software company** that operates through various **intelligent agents** holding different roles, including Chief Executive Officer , Chief Product Officer , Chief Technology Officer , programmer , reviewer , tester , art designer .

What is next?

Agents

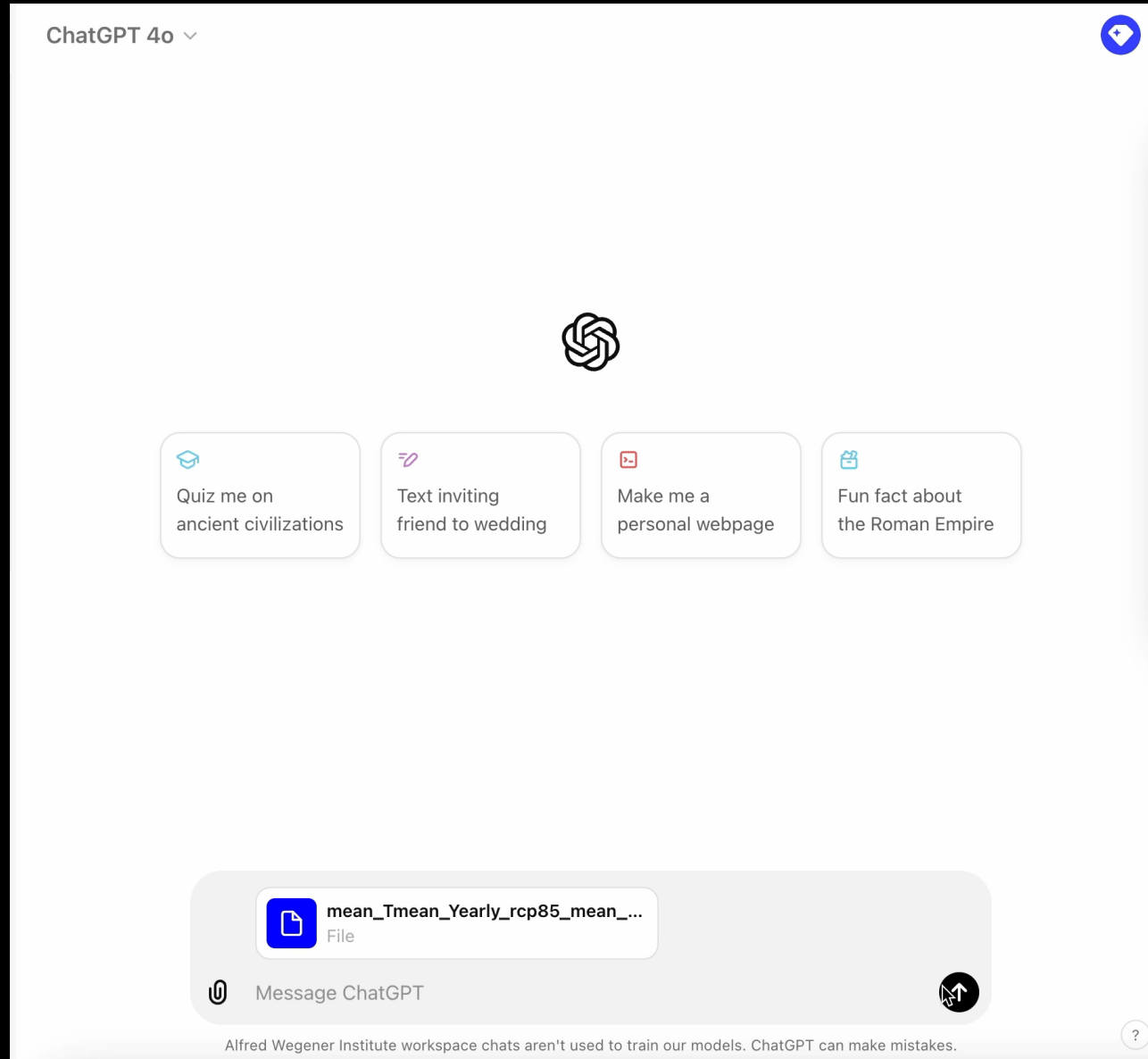
Create an assessment of climate change effects on building houses here.



ClimDev stands as a **virtual climate service company** that operates through various **intelligent agents** holding different roles, including Chief Scientist , Programmer , Climatologist , Economist , Geographer , Policy expert , art designer .

What is next?

Data processing



The screenshot shows the ChatGPT 4o interface. At the top left, it says "ChatGPT 4o" with a dropdown arrow. In the top right corner, there is a blue diamond icon. In the center, the OpenAI logo is displayed. Below the logo, there are four interactive buttons with icons and text: "Quiz me on ancient civilizations" (with a graduation cap icon), "Text inviting friend to wedding" (with a wedding ring icon), "Make me a personal webpage" (with a computer monitor icon), and "Fun fact about the Roman Empire" (with a gift icon). At the bottom, there is a message input area. It contains a file upload icon and a file name "mean_Tmean_Yearly_rcp85_mean_..." with the word "File" underneath. Below the input area, it says "Message ChatGPT" with a send button icon. At the very bottom, there is a small disclaimer: "Alfred Wegener Institute workspace chats aren't used to train our models. ChatGPT can make mistakes." and a help icon.

What is next?

- Work on the evaluation of LLM results – how to control quality, how to evaluate improvements?
- Work on communicating data uncertainty.
- Finetuned models in combination with RAG (can be smaller than general-purpose LLMs and run locally).
- Download and process data depending on the question.
- Take into account graphical information (maps) and information from surrounding points.
- Agents with individually tuned LLMs for each agent.
- Run on-demand impact models (e.g., AI-based) relative to the question.
- More data sources


Conclusions


- Current LLMs are great at summarizing data from different sources and putting it into the context of a question.
- Augmentation with data improves the quality of the response. More data is better, and local data is better.
- LLMs can provide a good interface to deliver climate and weather information at scale.
- We can not only ask “what is the weather” questions but “how will weather affect X” type of questions.


National weather services have a lot of information about local weather and geography and are well-positioned to create LLM-based interfaces to answer questions that connect human activities and weather/climate.

LLMs as interface for weather data

[Chat](#) [Weather Plot](#) [Weather Data](#)

 **weatherChat**

 weather informed conversations

 Let's talk about the weather! Give me a location to get started.

Please enter the location you are interested in for the weather forecast:

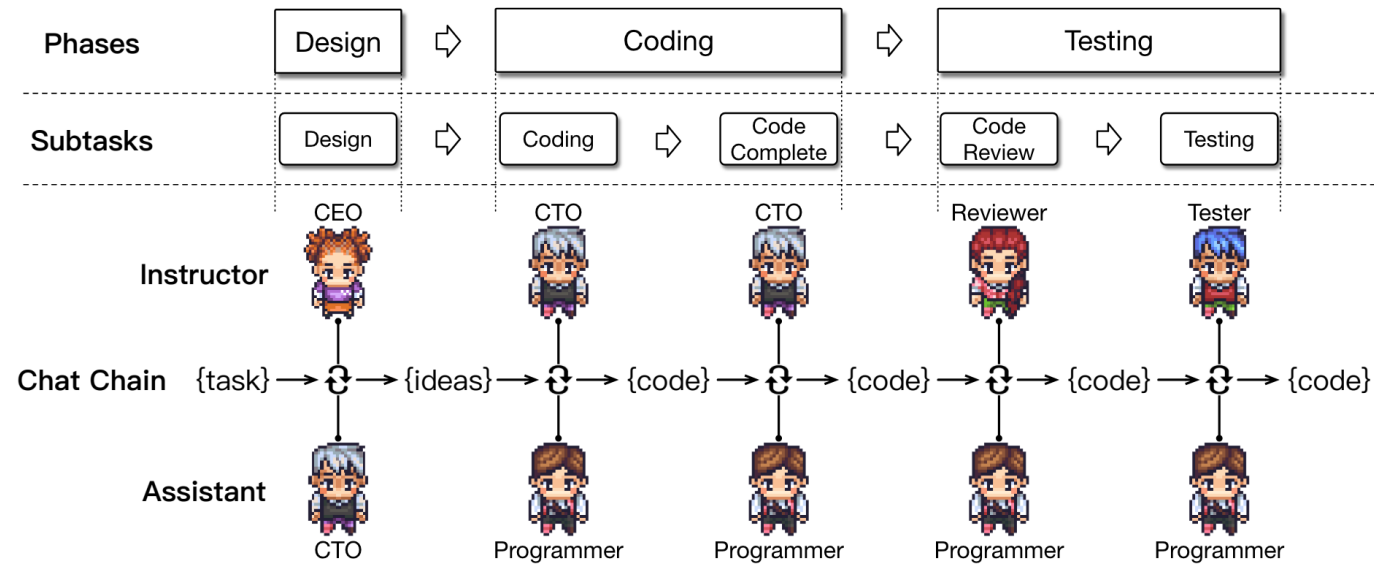


Figure 2: Upon receiving a preliminary task requirement (e.g., “develop a Gomoku game”), these software agents engage in multi-turn communication and perform instruction-following along a chain-structured workflow, collaborating to execute a series of subtasks autonomously to craft a comprehensive solution.